

PapillomaVirus  
Episteme, PaVE

# User Manual

The PapillomaVirus Episteme (PaVE) has been established to provide information and analytic resources to the scientific community for research on the Papillomaviridae family of viruses. This manual describes how to navigate to the data and analytical tools available in PaVE. The FAQs at the end of this manual provide detailed information about some of the functionalities and browser requirements.

The PaVE team includes members from the laboratory of Dr. Alison McBride at the Laboratory of Viral Diseases in the Division of Intramural research and the Bioinformatics and Computational Biosciences Branch (BCBB) headed by Dr. Yentram Huyen at Office of Cyber Infrastructure and Computational Biology (OCICB) in the NIAID/NIH.

# PaVE

## User Manual

### TABLE OF CONTENTS

#### 1. CONTENTS

<b>2. OVERVIEW .....</b>	<b>4</b>
<b>3. NAVIGATION IN PAVE .....</b>	<b>4</b>
3.1. Home Page .....	4
3.2. Top Navigation Bar .....	5
<b>4. SEARCH PAVE .....</b>	<b>6</b>
4.1. Keyword Search .....	6
Database Menu .....	7
Filtering Criteria .....	7
Search Results .....	8
4.2. PV Specific Blast Search .....	9
<b>5. ANALYZE .....</b>	<b>10</b>
5.1. Multiple Sequence Alignment of proteins .....	10
5.2. Protein structure .....	13
Table of Available protein structures. ....	13
Search for available pv protein structures .....	14
5.3. L1 typing tool .....	15
<b>6. EXPLORE .....</b>	<b>18</b>

6.1.	Browse reference genomes .....	19
6.2.	Variant Genomes .....	19
6.3.	Transcript maps .....	20
6.4.	Proteins .....	20
6.5.	Epigenetics .....	20
6.6.	Taxonomy .....	20
6.7.	Review Chapters.....	20
6.8.	Image viewer .....	20
6.9.	Movies .....	22
6.10.	Phylogenetic tree .....	22
7.	SEQUENCE AND STRUCTURE VIEWERS .....	23
7.1.	Locus viewer .....	23
7.2.	Structure Viewer .....	24
8.	FREQUENTLY ASKED QUESTIONS (FAQS).....	25

## 2. OVERVIEW

The PapillomaVirus Episteme (PaVE) was initiated in 2008 to provide information and bioinformatics resources to the scientific community for research on the Papillomaviridae. The goal of the PaVE was to provide an integrated resource for the storage and analysis of papillomavirus genome sequences and related information. The first release went live in April 2009. The PaVE is a freely accessible, web-based tool (<http://pave.niaid.nih.gov>) created around a relational database. Reference papillomavirus (PV) genome sequences are extracted from publicly available databases, corrected for sequence errors and uniformly re-annotated using a custom-created tool. The PaVE is frequently updated and includes 307 annotated PV genomes, with associated genes, regions, proteins and protein structures, which users can explore, analyze or download. The update also included annotation of additional proteins encoded by spliced transcripts. In addition, because of recent advances in Next-Gen sequencing, several putative novel genomes have been described (See [de Villiers et al., 2004](#) and [Bernard et al., 2010](#)) that do not meet all these requirements, and will therefore not be recognized as novel viral types by the [International Human Papillomavirus Reference Center](#). In order to reflect the known papillomavirus diversity PaVE has chosen to include viruses (14) that meet the “90% sequence identity” rule, even if they do not meet the other criteria. These viruses will be identified by the appendix “nr”. For more details refer to the [Taxonomy Concept](#) page.

The PaVE is a valuable resource for PV research; investigators can download complete sets of specific sequences for different viral proteins or regulatory regions, or can run PV specific BLAST searches against the PaVE databases. In 2011 we added transcript maps and a novel protein structure viewer. With the Structure Viewer, users can align selected protein sequences with the structure of a homologous protein, enabling easy identification of the location of conserved or divergent protein sequences. In 2012, we uniformly annotated the genomes and added *cis*-elements such as E2 binding sites. In 2013, we added an L1 typing tool, a Multiple Sequence Alignment module and an Image Viewer to allow the visualization of clinical and histological lesions associated with papillomavirus infection. It includes links to the 2013 [Special Issue of Virology](#), authored by experts in the fields providing detailed information and analysis for each viral protein, the URR, viral transcripts, and variants, as well as information about viral evolution, classification and disease. This issue is intended to be a reference source for the papillomavirus field. Papers that reference PaVE can be viewed [here](#).

## 3. NAVIGATION IN PAVE

### 3.1. HOME PAGE

The Home Page (Figure 1. PaVE Home) is one of two primary ways to use the PaVE website. The Home page provides quick links to the Search, Analyze, and Explore pages and other papillomavirus resources. It also includes overviews of the site and recent updates

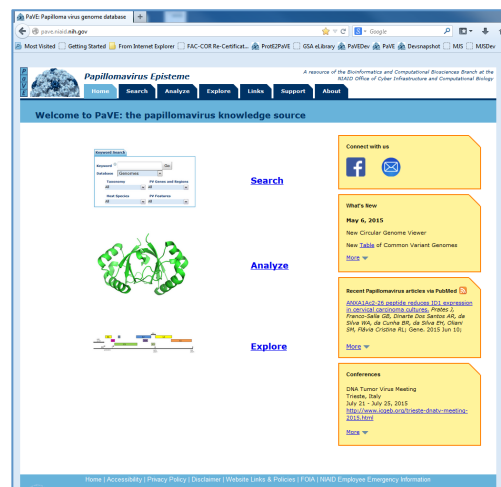


Figure 1. PaVE Home

## 3.2. TOP NAVIGATION BAR

The site can also be navigated through the tabs in the navigation bar (Figure 2. PaVE Navigation bar) that is visible at the top of all PaVE pages. The first tab links to the home page, while the other 6 tabs link to features listed below:



Figure 2. PaVE Navigation bar

1. **Home:** Links to the Home Page as described above
2. **Search:** PaVE provides the ability to search the database for sequences using **Keyword** Search or **PV Specific Blast** Search and for images on the Image Viewer.



Figure 3. Search for PV Sequences, Protein Structures and Images

3. **Analyze:** PaVE provides analysis tools for identification of potentially new PV type(s), multiple protein sequence alignments and for viewing papillomavirus-related protein structures from the Protein Data Bank (PDB).

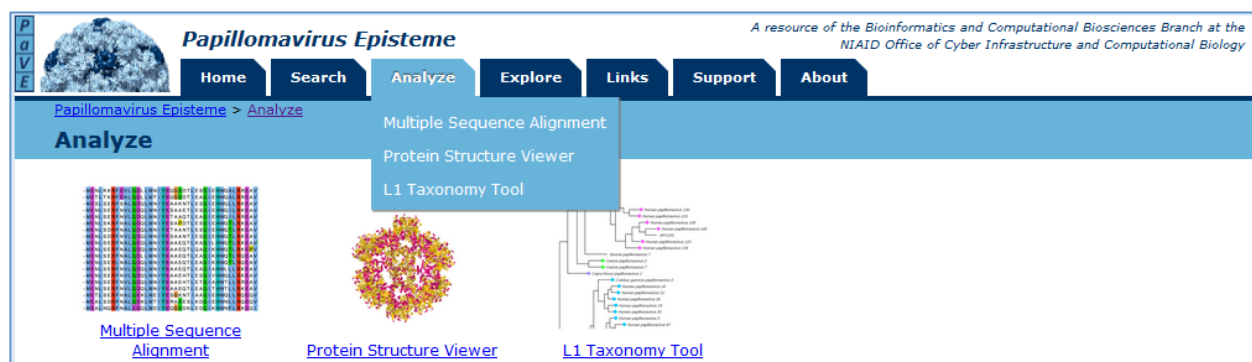


Figure 4. Analyze

4. **Explore:** PaVE provides tools to browse PV-related information included tabular and tree views of human and animal PaVE reference clones, taxonomic and classification concepts, transcript maps, proteins, images, and recent reviews focusing on functional genomics of PV.

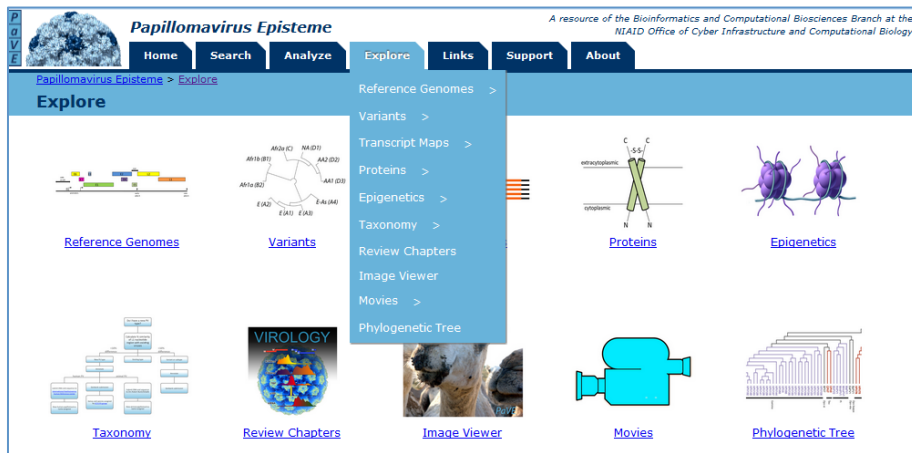


Figure 5. Explore

5. **Links:** Useful papillomavirus-related links
6. **Support:** A form to provide feedback, and a list of FAQs
7. **About:** Background and overview of the PaVE project

The Search, Analyze and Explore tabs are described in detail below.

## 4. SEARCH PAVE

### 4.1. KEYWORD SEARCH

The *Keyword Search* interface (figure 2) enables users to query the PaVE database with a simple keyword for

- reference genomes,
- genes and named regions,
- protein sequences
- protein structures

Search results can be filtered dynamically by

- taxonomic classification
- host species
- genes or regions
- reference/non-reference genomes

Figure 6. Keyword Search Interface

## DATABASE MENU

*Search Tools* features a *Database* menu (figure 7) that enables users to choose the type of data to be retrieved, namely, reference genome (*Genomes*), gene and named region (*Genes and regions*), protein sequences (*Protein sequences*) or protein structures (*Structures in PDB*).

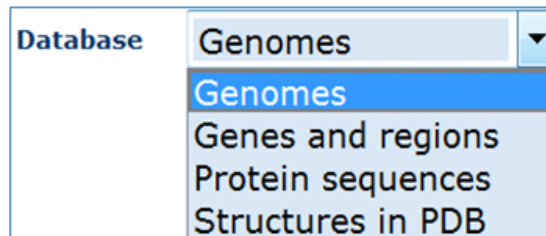


Figure 7. *Keyword Search Interface: Database menu.*

## FILTERING CRITERIA

1. *Non-reference genomes* can be included or excluded from the search results. Non-reference viruses (designated by 'nr') in PaVE are those with L1 sequences that do not share more than 90% nucleotide sequence identity with their closest neighbor, which however, do not meet other criteria as specified in [de Villiers et al., 2004](#) and [Bernard et al., 2010](#) in order to be recognized as new types. For more details, refer to **the section [Taxonomy concept - Non-reference genomes](#)** on PaVE.
2. The *Taxonomy* filter enables users to filter at the genus or species level (figure 8a). Clicking the taxonomy box provides a dialogue box to enable users to select at the level of the genus or expand the genus and select at the species level (click on the + to display the corresponding species).
3. The *Host Species* filter (figure 8b) enables users to filter by the scientific name of host animal from which the virus was isolated.
4. The *PV Genes and Regions* filter (figure 8c) enables users to select sequences that contain the selected genes or specific regions such as the Upstream Regulatory Region (URR).

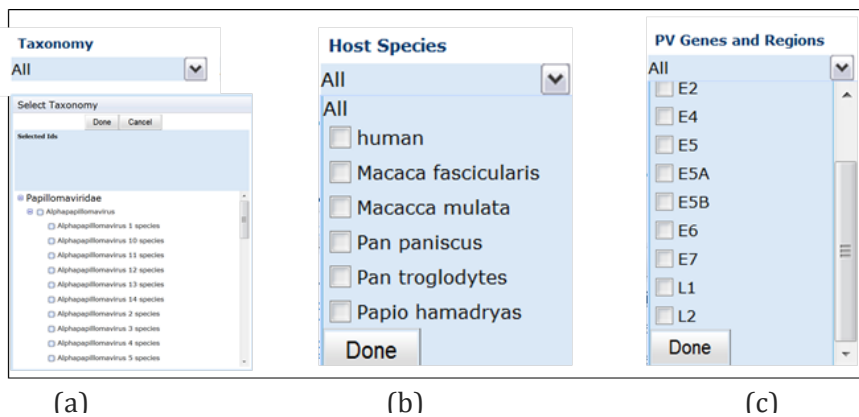


Figure 8. *Search Tools interface: Filtering criteria.*

## SEARCH RESULTS

Search results are displayed below the search interface in a table and include the sequence name and a description of the sequence (figure 9). Users can download the DNA and protein sequences for further analyses in appropriate formats (for example, FastA or GenBank format) by either selecting specific sequences or clicking “Select All” and clicking on the “Download” buttons for the specific format (FastA, GenBank). Clicking on the sequence name will render a graphical representation of the sequence together with annotations (see figure 22a, Locus Viewer).

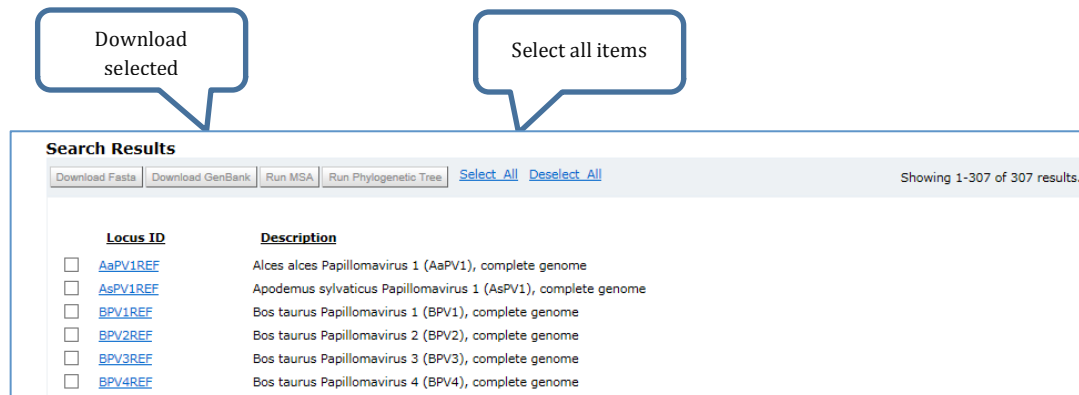


Figure 9. Search Results view.

Search results for protein sequences (figure 10) will also include a column to indicate the availability of a PDB structure (red filled circles) of this protein or a related structure (orange filled circles) of a homologous protein (with greater than 10% amino acid sequence identity). Clicking on these circles will display the structure and sequence alignments with homologous proteins (see figure 24, Structure viewer).

Selected protein sequences can be further analyzed:

1. Aligned using the “Run MSA” button
2. Aligned and depicted within a phylogenetic tree to show relatedness between sequences using the “Run Phylogenetics Tree” button





## 5. ANALYZE

PaVE provides analysis tools (figure 12) for identification of potentially new PV type(s), multiple protein sequence alignments and for viewing papillomavirus-related protein structures from the Protein Data Bank (PDB).

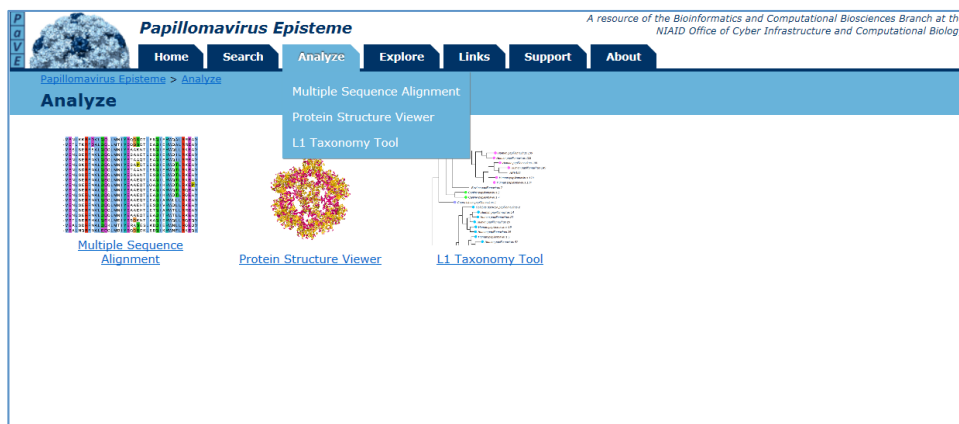


Figure 12. Analyze Interface

### 5.1. MULTIPLE SEQUENCE ALIGNMENT OF PROTEINS

Users can perform multiple sequence alignments (MSA) of proteins using the [COBALT](#) (Constraint-based Multiple Protein Alignment Tool). The Alignment tab on the top bar links to a form that enables users to submit their FastA formatted query sequences either by pasting them in the window or by uploading a file containing these sequences and clicking Analyze (figure 13a). Users can also perform MSA of PaVE sequences by selecting them from the Search Page and clicking on the “Run MSA” button (figure 13b).

Upon performing MSA, users can either view the alignment in a CLUSTAL alignment view (default, figure 13d) or within [JalView](#)\* (figure 13e), an external program for multiple sequence alignment editing, visualization and analysis. Jalview has been integrated into PaVE so that it runs in the user’s browser (not compatible with Firefox Version 36 and higher). You can obtain more information about the JalView program at <http://www.jalview.org/about/Documentation>.

Two links on the default CLUSTAL view page “[Download MSA Fasta Result](#)” and “Download MSA EPS Image” enables users to download the sequences in mFastA format (with the original IDs) or the alignment image generated in JalView.

\*Note: A popup Window may request the permission of users to run the JalView alignment viewer program (figure 13c). Users need to check, “I accept the risk” to run this program in the browser. In the event that you get an error message that your java version needs to be updated or that you need to enable java in your browser, please update it or this tool will not work.

**Papillomavirus Episteme**  
A resource of the Bioinformatics and Computational Biosciences Branch at the  
NIAID Office of Cyber Infrastructure and Computational Biology

Home Search Analyze Explore Links Support About

Papillomavirus Episteme > Analyze > Multiple Sequence Alignment

## Multiple Sequence Alignment

Protein Structure Viewer

MSA Uses the NCBI tool [COBALT\(Conserved Domain Architecture Annotation Tool\)](#) which computes the alignment using conserved domain and local sequence similarity information.

Alignment can be visualized by [Jalview](#).

Query sequences can be input below or by selecting protein sequences on the [Search page](#) and running MSA.

☐ Paste into the form below  
☐ Upload file

Input Sequence (FASTA format)  
At least two input protein sequences needed ⓘ

Home | Accessibility | Privacy Policy | Disclaimer | Website Links & Policies | FOIA | NIAID Employee Emergency Information

NIH National Institute of Allergy and Infectious Diseases USA.gov

Figure 13a. Multiple Sequence Alignment Interface.

## Search Database

Keyword ⓘ  Go

Database  ☐ Include [Non-reference genomes](#)

▼ Filter by

Dynamic search: results are automatically updated with the filter applied. Please do not refresh the page if the taxonomy dropdown does not load immediately, it may take a minute to load.

Taxonomy Click to see selected ▼ PV Genes and Regions E1 ▼

Host Species All ▼ PV Features All ▼

**Selected Filters:** [Clear All](#)

Database: [Protein sequences]

PV Genes and Regions: [E1] X

Taxonomy: [Alphapapillomavirus] X

includeNR: [false] X

## Search Results

[Select All](#) [Deselect All](#) Showing 1-77 of 77 results.

Locus ID ▲	Associated/Homologous Structure ⓘ	Description
<input type="checkbox"/> <a href="#">CgPV1-E1</a>		Colobus guereza Papillomavirus 1 (CgPV1), E1 protein
<input type="checkbox"/> <a href="#">HPV2-E1</a>		Human papillomavirus 2 (HPV2), E1 protein
<input type="checkbox"/> <a href="#">HPV3-E1</a>		Human papillomavirus 3 (HPV3), E1 protein
<input type="checkbox"/> <a href="#">HPV6-E1</a>		Human papillomavirus 6 (HPV6), E1 protein
<input type="checkbox"/> <a href="#">HPV7-E1</a>		Human papillomavirus 7 (HPV7), E1 protein
<input type="checkbox"/> <a href="#">HPV10-E1</a>		Human papillomavirus 10 (HPV10), E1 protein
<input type="checkbox"/> <a href="#">HPV11-E1</a>		Human papillomavirus 11 (HPV11), E1 protein

Figure 13b. Multiple Sequence Alignment via the Search Page.

**MSA Analysis Result**

[Download MSA Mfasta Result](#)  
[Download MSA EPS Image](#)

Launch Jalview with Full Sequence IDs:

```

CLUSTAL
1cl|1/1-612   MAD--NKGTE-----DWFLVEA--TDCEET--LEETSLGDLDNVSCVSDLSDLLDEA
1cl|2/1-643   MED--SEGTDGTEEDGCRAGGFHVEAIITHGQR-Q-VSSDEDEDETETGEDL-DFIDNR
1cl|3/1-659   MDD--TSGTGECSELERAGGFHVEAIVDRRTG-DTVSSDEDEEE-DGGEDLVDFIDDR
1cl|4/1-599   MAD--KGTDN-----FDLEGNNWYIVHEAECTDSIDTLDLDCDESND--SNISNLIDDD
1cl|5/1-606   MTDPNKSGSTS----KEGFGDWCLL--EADCDV--ENDLQGLFERD--SDISDLDDT
1cl|6/1-649   MAD--DSGTEN-----EGSGCTGWFHVEAIVQHPGTG-IQISDDEDEEVEDSGYDMVDFIDDS

1cl|1/1-612   PQSQGNSLE---LFHQQSELESEQELNALKRKLKLYSPQARSAD---ETDIASISPRLETI
1cl|2/1-643   VPGDGGEIPLQ-LYAQQTADDEATVQALKRKFVASPLSACSC----IENDLSPRLDIAI
1cl|3/1-659   VPGDGGEVAQE-LLQQAADDVVEVQTVKRKFAPSPYFVPC-VHPSIENELSPRLDAI
1cl|4/1-599   VVDQGNLSLA---LYNAQINEDCDNALAHLKRYNKS-----EQAVALESPQLQAV
1cl|5/1-606   ELEQGNLSLE---LFHQQCEQSEELQKLKRYLSP-----KAVAQLSPRLESI
1cl|6/1-649   -NITHNSLEAQLFNRQEADTHYATVQDLKRYLGSPPVSPINTIAEAVESEISPRLDIAI

1cl|1/1-612   SITKQDKKRYRQLFSQ---DDSGLELQLD-----GDEESQGG
1cl|2/1-643   SLNRKSEK-AKRRLFET-EPPDSGYGNTQMVVGT-PEEVT-----GDEESQGG
1cl|3/1-659   KLGRQTSK-AKRRLFEL---PDSGYGNTQVDTESGPKVQVDICKTSQQDGCQGADEGRGR
1cl|4/1-599   KITPERHS--KRRLF-----QDSGIFEAE-----GTQVEKHG
1cl|5/1-606   SLSPQQKS--KRRLFAE---QDSGLETLNN-----GTQVEKHG
1cl|6/1-649   KLTRQPKK-VKRRLFQTRELTDSGYGYSEVAGT-----GTQVEKHG

1cl|1/1-612   -----ETEN-IDESTQVD-QQQKEHTGEVGAAGV---DILKASNIR
1cl|2/1-643   RP--VED--QEEERQGGD-----GEADLTIV-HTPQSGTDAAGSV-L---TLRSSNLK
1cl|3/1-659   NV--GGNGSQEEERAGGD-----GEESQT--ESVQDTTACG-V-L---AILKASNHK
1cl|4/1-599   -----NS---LTQVESES-QAGPSSQDGGGIDNL---LLQSSNRR
1cl|5/1-606   -----EADVTPEVEVPAIDSRPDDEGGSGDVIHYTALLRSSNKK
1cl|6/1-649   VPENGDDG--QEKDTRDIEGEE-HTEAEAPT-NSVREHAGTAGI--L---ELLKCKDLR

1cl|1/1-612   AALLSRFKDTAGVSFTDLIRSYKSNKTCGDDWVLAVGVRENLIDSVKELLQTHCVYIQL
1cl|2/1-643   ATLLSKFKDLFGVGFYELVRQFKSSKTACADWVVCAYGVYVAVAEGLKLIQPHQYAI
1cl|3/1-659   ATLLGKFKEQFGLGFNELLIRHFKSNKTVCSDDWVVCVGVYCTLAESFKTLIQPCYAI
1cl|4/1-599   ATMLAKFKEWYGVSYNEITRIYKSDKSCSDNWVIVIFRAAVEVLESSKIVLQHCYIYQV
1cl|5/1-606   ATMLAKFKEFVGFGFNLTRQFKSHKTCCKDWVSVYAVHDDLFESSKQLLQHCYIYQV
1cl|6/1-649   AALLGKFKECFGLSFIDLIRPFKSDKTTCLDWVAVFGIHHISIEAFQKLEPLSLYAI

1cl|1/1-612   EHAVTEKNRFLFLVRFKAQKSRQETVIKLTITLIPVDASYILSEPPKSRVAAALFWYKR
1cl|2/1-643   QVQTSSWGVVFMLLRYNCAKNRDSVSKNMSMLNIPKHMILPEPKLRSTPAALYWKYK
1cl|3/1-659   QVQTSSWGVVFMLLRYNCAKNRDSVSKNMSMLNIPKHMILPEPKLRSTPAALYWKYK

```

Figure 13c. CLUSTAL output of alignment.

**MSA Analysis Result**

[Download MSA Mfasta Result](#)  
[Download MSA EPS Image](#)

Launch Jalview with Full Sequence IDs:

```

CLUSTAL
1cl|1/1-612   MAD--NKGTE-----DWFLVEA--TDCEET--LEETSLGDLDNVSCVSDLSDLLDEA
1cl|2/1-643   MED--SEGTDGTEEDGCRAGGFHVEAIITHGQR-Q-VSSDEDEDETETGEDL-DFIDNR
1cl|3/1-659   MDD--TSGTGECSELERAGGFHVEAIVDRRTG-DTVSSDEDEEE-DGGEDLVDFIDDR
1cl|4/1-599   MAD--KGTDN-----FDLEGNNWYIVHEAECTDSIDTLDLDCDESND--SNISNLIDDD
1cl|5/1-606   MTDPNKSGSTS----KEGFGDWCLL--EADCDV--ENDLQGLFERD--SDISDLDDT
1cl|6/1-649   MAD--DSGTEN-----EGSGCTGWFHVEAIVQHPGTG-IQISDDEDEEVEDSGYDMVDFIDDS

1cl|1/1-612   PQSQGNSLE---LFHQQSELESEQELNALKRKLKLYSPQARSAD---ETDIASISPRLETI
1cl|2/1-643   VPGDGGEIPLQ-LYAQQTADDEATVQALKRKFVASPLSACSC----IENDLSPRLDIAI
1cl|3/1-659   VPGDGGEVAQE-LLQQAADDVVEVQTVKRKFAPSPYFVPC-VHPSIENELSPRLDAI
1cl|4/1-599   VVDQGNLSLA---LYNAQINEDCDNALAHLKRYNKS-----EQAVALESPQLQAV
1cl|5/1-606   ELEQGNLSLE---LFHQQCEQSEELQKLKRYLSP-----KAVAQLSPRLESI
1cl|6/1-649   -NITHNSLEAQLFNRQEADTHYATVQDLKRYLGSPPVSPINTIAEAVESEISPRLDIAI

1cl|1/1-612   SITKQDKKRYRQLFSQ---DDSGLELQLD-----GDEESQGG
1cl|2/1-643   SLNRKSEK-AKRRLFET-EPPDSGYGNTQMVVGT-PEEVT-----GDEESQGG
1cl|3/1-659   KLGRQTSK-AKRRLFEL---PDSGYGNTQVDTESGPKVQVDICKTSQQDGCQGADEGRGR
1cl|4/1-599   KITPERHS--KRRLF-----QDSGIFEAE-----GTQVEKHG
1cl|5/1-606   SLSPQQKS--KRRLFAE---QDSGLETLNN-----GTQVEKHG
1cl|6/1-649   KLTRQPKK-VKRRLFQTRELTDSGYGYSEVAGT-----GTQVEKHG

1cl|1/1-612   -----ETEN-IDESTQVD-QQQKEHTGEVGAAGV---DILKASNIR
1cl|2/1-643   RP--VED--QEEERQGGD-----GEADLTIV-HTPQSGTDAAGSV-L---TLRSSNLK
1cl|3/1-659   NV--GGNGSQEEERAGGD-----GEESQT--ESVQDTTACG-V-L---AILKASNHK
1cl|4/1-599   -----NS---LTQVESES-QAGPSSQDGGGIDNL---LLQSSNRR
1cl|5/1-606   -----EADVTPEVEVPAIDSRPDDEGGSGDVIHYTALLRSSNKK
1cl|6/1-649   VPENGDDG--QEKDTRDIEGEE-HTEAEAPT-NSVREHAGTAGI--L---ELLKCKDLR

1cl|1/1-612   AALLSRFKDTAGVSFTDLIRSYKSNKTCGDDWVLAVGVRENLIDSVKELLQTHCVYIQL
1cl|2/1-643   ATLLSKFKDLFGVGFYELVRQFKSSKTACADWVVCAYGVYVAVAEGLKLIQPHQYAI
1cl|3/1-659   ATLLGKFKEQFGLGFNELLIRHFKSNKTVCSDDWVVCVGVYCTLAESFKTLIQPCYAI
1cl|4/1-599   ATMLAKFKEWYGVSYNEITRIYKSDKSCSDNWVIVIFRAAVEVLESSKIVLQHCYIYQV
1cl|5/1-606   ATMLAKFKEFVGFGFNLTRQFKSHKTCCKDWVSVYAVHDDLFESSKQLLQHCYIYQV
1cl|6/1-649   AALLGKFKECFGLSFIDLIRPFKSDKTTCLDWVAVFGIHHISIEAFQKLEPLSLYAI

1cl|1/1-612   EHAVTEKNRFLFLVRFKAQKSRQETVIKLTITLIPVDASYILSEPPKSRVAAALFWYKR
1cl|2/1-643   QVQTSSWGVVFMLLRYNCAKNRDSVSKNMSMLNIPKHMILPEPKLRSTPAALYWKYK

```

Start Jalview

View alignment in Jalview

Figure 13d. Option to view alignment in JalView.

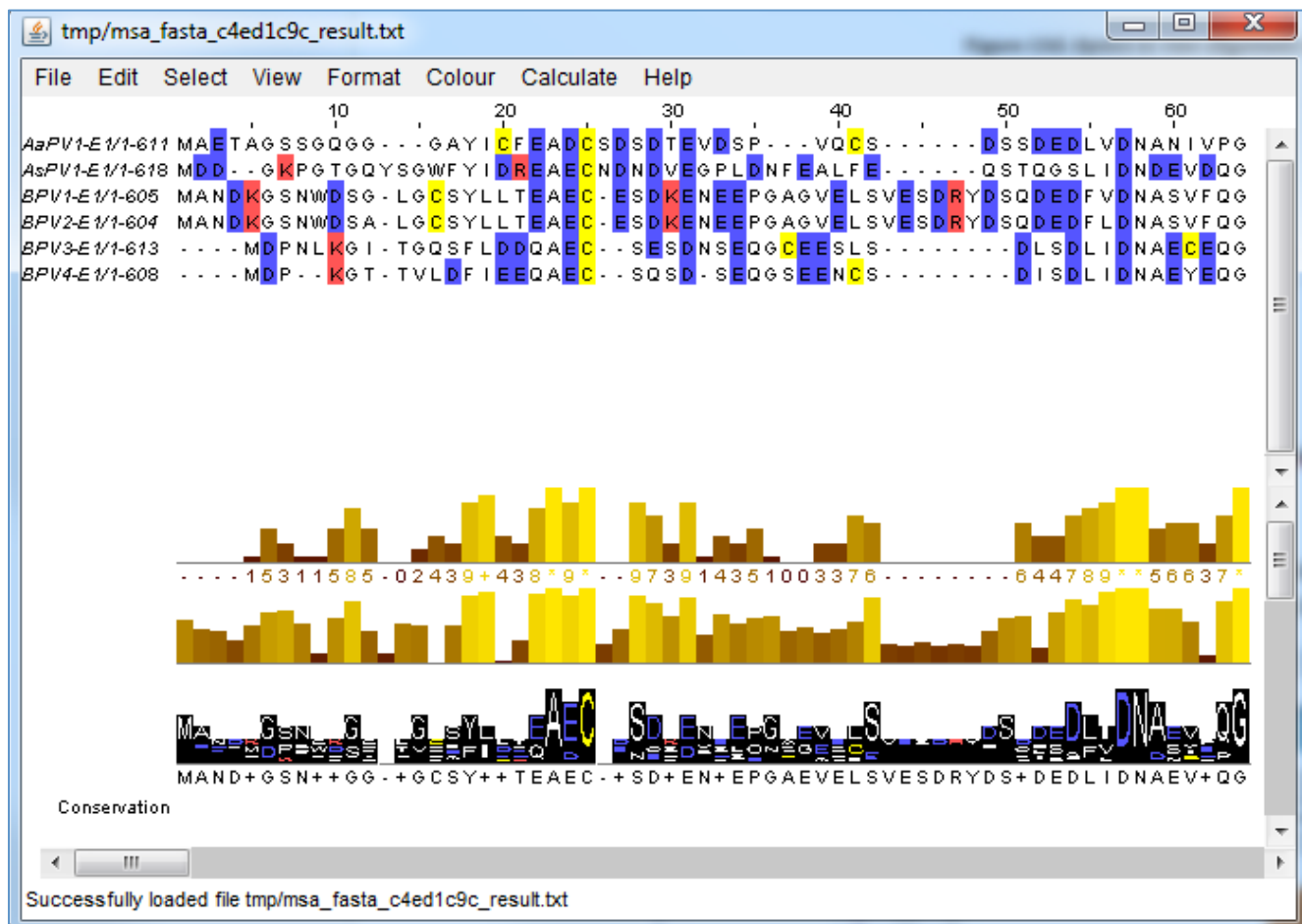


Figure 13e. Option to view Alignment in JalView.

## 5.2. PROTEIN STRUCTURE

The Protein structure interface (figure 14) enables users to either browse the Table of available structure or Search for available PV protein structures in PaVE via the Search Interface. (figure 16)

**Papillomavirus Episteme**  
A resource of the Bioinformatics and Computational Biosciences Branch at the  
NIAID Office of Cyber Infrastructure and Computational Biology

Home Search Analyze Explore Links Support About

Papillomavirus Episteme > Analyze > Protein Structure Viewer

The Protein Structure Viewer module displays the protein structures of papillomavirus proteins. The structure viewer can be accessed in several ways. A list of available structures is provided, and users can retrieve proteins with associated structures in either the **Protein Sequences** or **Structures in PDB** databases. Proteins with an associated structure are indicated by a light orange icon. Clicking this icon will display this protein in the **Structure Viewer**. In the **Protein Sequences** database, many proteins have a homologous structure indicated by a light orange icon. Accessing the Structure Viewer through this icon will display the known structure of a homologous protein and a pairwise alignment of both proteins. This allows users to identify regions of sequence similarity or divergence on the protein structure.

Protein	PDB ID	Structure	Accession	Model
1	1A22	1A22	1A22	1A22
2	1A23	1A23	1A23	1A23
3	1A24	1A24	1A24	1A24
4	1A25	1A25	1A25	1A25
5	1A26	1A26	1A26	1A26
6	1A27	1A27	1A27	1A27
7	1A28	1A28	1A28	1A28
8	1A29	1A29	1A29	1A29
9	1A30	1A30	1A30	1A30
10	1A31	1A31	1A31	1A31

[Table of available PV protein structures](#)

[Search for available PV protein structures](#)

Figure 14. Protein Structure Interface.

TABLE OF AVAILABLE PROTEIN STRUCTURES.

Clicking on the PDB ID will display the structure and sequence alignments with homologous proteins (see figure 24, structure viewer)

Table of Available Protein Structures			
Protein	PDB ID	Domain and/or complex	Virus
E1	<a href="#">1F08</a>	DBD	BPV1
	<a href="#">1K5X</a>	DBD + DNA	BPV1
	<a href="#">1K5Y</a>	DBD + DNA	BPV1
	<a href="#">1R9W</a>	DBD	HPV18
	<a href="#">1TUE</a>	E1 + E2	HPV18
	<a href="#">2GXA</a>	hexamer + DNA	BPV1
E2	<a href="#">2V9P</a>	hexamer	BPV1
	<a href="#">1A7G</a>	DBD	HPV31
	<a href="#">1BY9 *</a>	DBD	HPV16
	<a href="#">1DBD *</a>	DBD	BPV1
	<a href="#">1DHM *</a>	DBD	HPV31
	<a href="#">1DTQ</a>	TA	HPV16
	<a href="#">1F9F</a>	DBD	HPV18
	<a href="#">1JJ4</a>	DBD + DNA	HPV18
	<a href="#">1QQH *</a>	TA	HPV18
	<a href="#">1JJH</a>	DBD	BPV1
	<a href="#">1R6K</a>	TA	HPV11
	<a href="#">1R6N</a>	TA	HPV11
	<a href="#">1R8H *</a>	DBD	HPV6
	<a href="#">1R8P *</a>	DBD	HPV16
	<a href="#">1TUE</a>	E1 + E2	HPV18
	<a href="#">1ZZF *</a>	DBD	HPV16

Figure 15. Table of Available Protein Structures

#### SEARCH FOR AVAILABLE PV PROTEIN STRUCTURES

Clicking on the orange filled circles will display the structure and sequence alignments with homologous proteins (see figure 24, structure viewer).

### Search Database

Keyword

Go

Database

Structures in PDB

☐ Include [Non-reference genomes](#)

Filter by

Taxonomy

All

Host Species

All

PV Genes and Regions

All

PV Features

All

Dynamic search: results are automatically updated with the filter applied. Please do not refresh the page if the taxonomy dropdown does not load immediately, it may take a minute to load.

Selected Filters: [Clear All](#)

Database: [Structures in PDB]

includeNR: [false]

### Search Results

Download Fasta

Download GenBank

Run MSA

Run Phylogenetic Tree

[Select All](#)

[Deselect All](#)

Showing 1-48 of 48 results.

Locus ID	Associated Structure	Description
<input type="checkbox"/> <a href="#">1A7G</a>		The Crystal Structure Of The E2 Dna-Binding Domain From Human Papillomavirus At 2.4 Angstroms.
<input type="checkbox"/> <a href="#">1BY9</a>		Crystal Structure Of The E2 Dna-Binding Domain From Human Papillomavirus Type-16: Implications For Its Dna Binding- Site Selection Mechanism.
<input type="checkbox"/> <a href="#">1DBD</a>		E2 Dna-Binding Domain From Papillomavirus Bpv-1.
<input type="checkbox"/> <a href="#">1DHM</a>		Dna-Binding Domain Of E2 From Human Papillomavirus-31, Nmr, Minimized Average Structure.
<input type="checkbox"/> <a href="#">1DTQ</a>		Crystal Structure Of The Complete Transactivation Domain Of E2 Protein From The Human Papillomavirus Type 16.

Figure 16. Search for protein structures

14

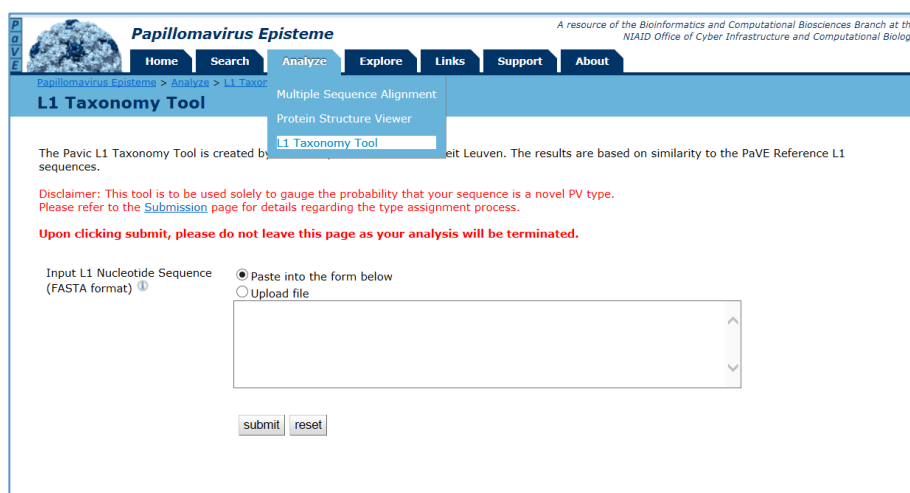
### 5.3. L1 TYPING TOOL

The L1 typing tool is designed to assist users in gauging the probability that their sequence belongs to a novel PV type based on the similarity of its L1 sequence to current PV types in the PaVE database. The tool was developed by Piet Maes, at the Katholieke Universiteit Leuven.

Users will need to submit one or more L1 nucleotide query sequence(s) in FASTA format either by pasting in the window or by uploading a file containing the sequence(s) and clicking “Submit” (Figure 17a). The resulting window will display a textual summary of the results (figure 17b). Additional reports of the results include results in txt and .csv formats, phylogenetic tree (PDF), and the newick tree (figures 17 c-d).

Note: \*If a single query sequence is used, it does not need to be in fastA format. For multiple query sequences, the L1 typing tool only accepts files that strictly adhere to the FASTA format. A sequence in FASTA format begins with a greater-than (“>”) symbol on the first line followed by sequence name or information. The L1 typing tool requires that a sequence name is entered after > without a space (eg >test). Sequence data begins on the second line.

Users should refer to the [Submission](#) page for details regarding the official type assignment process.



The screenshot shows the 'L1 Taxonomy Tool' interface. At the top, there's a navigation bar with links: Home, Search, Analyze, Explore, Links, Support, and About. Below this, a breadcrumb trail reads 'Papillomavirus Episteme > Analyze > L1 Taxonomy'. The main heading is 'L1 Taxonomy Tool'. A sub-header indicates 'Multiple Sequence Alignment' and 'Protein Structure Viewer'. A text box states: 'The PAVIC L1 Taxonomy Tool is created by [redacted] at Katholieke Universiteit Leuven. The results are based on similarity to the PaVE Reference L1 sequences.' Below this, a disclaimer in red text says: 'Disclaimer: This tool is to be used solely to gauge the probability that your sequence is a novel PV type. Please refer to the [Submission](#) page for details regarding the type assignment process.' A warning in red text follows: 'Upon clicking submit, please do not leave this page as your analysis will be terminated.' The input section is titled 'Input L1 Nucleotide Sequence (FASTA format)' and has two radio buttons: 'Paste into the form below' (selected) and 'Upload file'. Below the radio buttons is a large text area for pasting the sequence. At the bottom of the text area are 'submit' and 'reset' buttons.

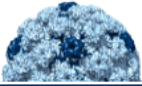
Figure 17a. L1 Typing Interface.

P

a

v

e



Papillomavirus Episteme

Home

Search

Analyze

Explore

Links

Support

About

A resource of the Bioinformatics and Computational Biosciences Branch at the NIAID Office of Cyber Infrastructure and Computational Biology

Papillomavirus Episteme

>

Analyze

>

L1 Taxonomy Tool

>

L1 Taxonomy Tool Analysis Results

L1 Taxonomy Tool Analysis Results

\* 1 sequence(s) found in query

\* Sequence 1: gi|333031.L1|cl|HPV16-L1.1|-Human-papillomavirus-16-(HPV16),-L1-gene

\* The query sequence is most similar to: HPV16 with 100.0% identity.

These results are to be used solely to gauge the probability that your sequence is a novel PV type. Please refer to the [Submission](#) page for details regarding the type assignment process.

Please save your results files after being displayed if you wish to save them since PAVE does not store these files.

Display Result

--Please select--

Text Summary

CSV Summary

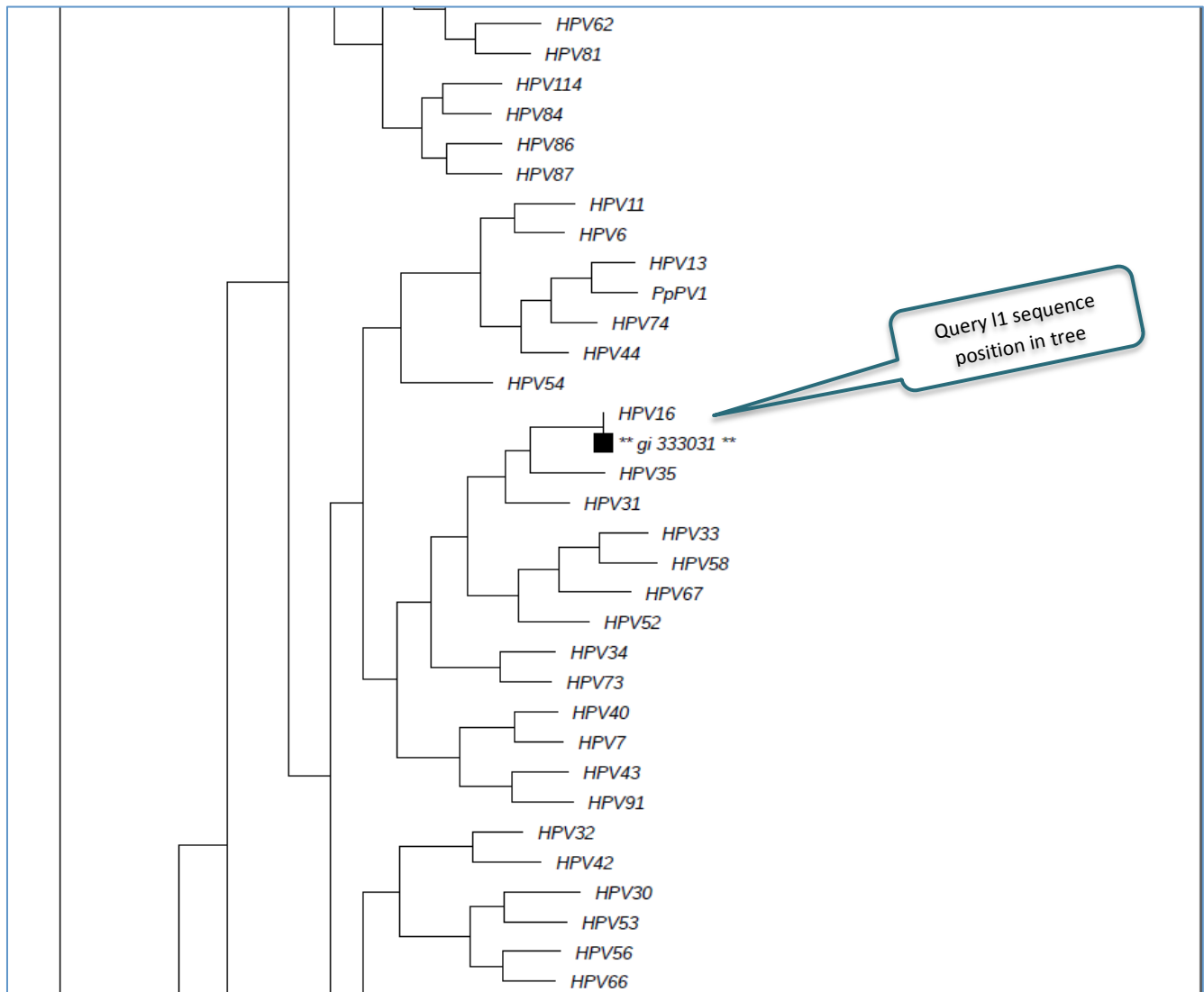
PDF Tree

Newick Tree

Figure 17b. L1 Typing – Summary Results.

16





**Figure 17c. L1 Typing - Phylogenetic Tree.**

```

This file is automatically generated by:
Automated Genotyping Tool for papillomaviruses - PaviC v1.1b
(c) 2013 Piet Maes - Clinical Virology, Rega Institute, Katholieke Universiteit Leuven, Belgium.
Reference strains, last update: 23-04-2012

* 1 sequence(s) found in query
-----

* Sequence 1: L1
Length: 1509 bps

Sequence:
ATGGGTGTCGTGGTTACCGGCGAGAATAAGTTCATCTTCTCTCCCGAGCCCATCACTAGAACTCTGTCCACTGATGAATATGTAACCGA
ACCAATCTCTTCTACCATGCAACATCTGAACGCTCTACTGCTGGTGGGACATCCTTTGTTTGGAGATCTCCAGTAATCAAAGTAACTATA
CCAAAAGTGTACCAAAATGCAATTTAGAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
CCAGAAACAGAAAGATTAGTTTGGGGCCTAAGAGGGATAGAGATAGTAGTAGGCGCCCTTTAGGTATTAGGAATAACGGGCCACCCCTCTT
TTAAATAAGTTAGATGATGCGAGAAATCCAAACAAATTATTAATACTCATGCAAAATGGAGATTCTAGACAAATACTGCTTTTGTATGCA
AAACAGACACAAATGTTCTCGTGGGCTGTACTCTCTGCTTCAAGTGAACACTGGACAAAGTAGTCGTTGCCAGGGGAAACAAAGTGAACCTT
GGGGACTGCCCGAGGGTGAATGATAGAGTCTGTCTAGAAAGATGGTACATGGATATTGGTTTGGGGCTATGGATTGTGCTGCT
TTACAGCAAGCAAGCTGTGATGCTCTTATAGTGTGTTCAAGCAACATGCAAAATATCCTGATTATATCAGAAATGAACCAAGGCTAT
GGCAACTCTATGTTTTTTTTTGCACGTGCGAGCAAAATGTATACAGGCACTTTTTTACTGCGGGGGGTTGGTGGGTGATAGGAGGCA
GTCCCAAAAGCCTGTATTTAAGCAGATGCTGAACCAAGCAACTTTAGCAACAAATAATGATAGGACACCAAGTGGCTCTATG
GTTTCATCTGATGCTCAATTGTTTAATAGATCTTACTGGCTTCAGCGATGTCAAGGCCAGAAATAATGGCAATTTGCTGGAGAAACAGTTA
TTTATTACAGTTGGAGATAATACAGAGGAACAAAGTTTATCTATCAGTATGAAAAACAAATGCAAGTACTACATATTCCAATGCTAAATTT
AATGATTTTCTAAGACATACTGAAGAATTTGATCTTTCTTTTATAGTTCAGCTTTGTAAAGTAAAGTTAACTCCGAAAAATCTAGCTAC
ATTCATACAAATGGACCTTAATTTTGAAGGATTGGCACTATCTGTATCTCAACCACTACCAATCTCTAGAAAGTCAATATAGGTTT
TTAGSGTCTTCTTGGCAGCAAAATGTCCAGAACAGGCGCTCTGAGCCCCAGACTGATCCTTATAGTCAATATAAATTTCTGGGAAGTC
GATCTCAGAGAAAGATGTCGGAACAATTAGACCAATTTCCACTAGGAAGGAAATTTCTATATCAAGTGGCATGACACCAAGCTACTGCT
ACTAGTTCACCAACCAAGCGCAAAACAGTGGTGTATCTACGTCAGCCAGCGCAGCGGTAAGGCTTAG

```

\* Origin of query sequence:  
\* The query sequence is most similar to: HPV1 with 100.0% identity.  
\* Conclusion:  
The sequence belongs to the Mupapillomavirus genus.

Figure 17d. L1 Typing Results –Text File.

## 6. EXPLORE

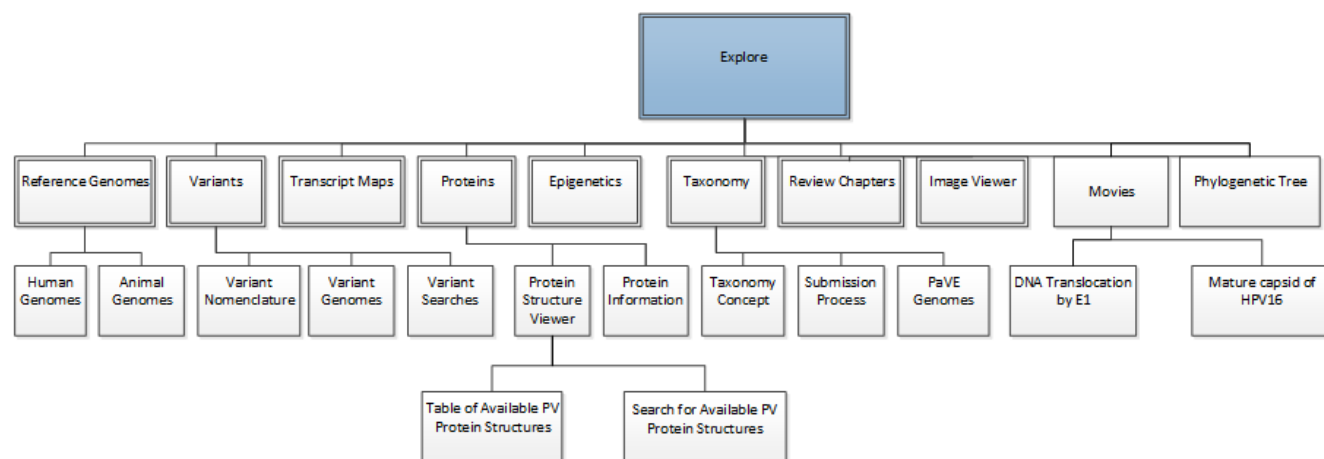


Figure 18a. Schematic View of the Explore tab menu choices

PaVE provides tools to browse PV-related information included tabular and tree views of human and animal PaVE reference clones, variants, taxonomic and classification concepts, transcript maps, proteins, images, movies of proteins in action and recent reviews focusing on functional genomics of PV (figures 18a and 18b).

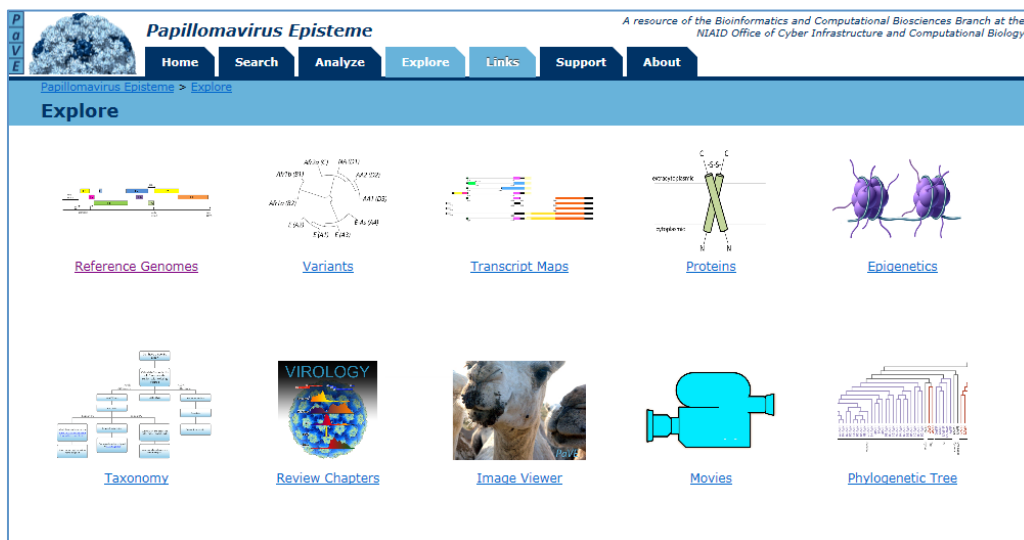


Figure 18b. *Explore* interface

## 6.1. BROWSE REFERENCE GENOMES

The *Reference Genomes* tab under the *Explore* tab links to the section of the website which includes detailed information about the Reference Genomes enabling users to browse all PaVE viral types in a zoomable phylogenetic tree (*Tree* tab, figure 10; inferred from the L1 sequences of PaVE Reference Clones) or via a tabular view of *Human* or *Animal* viral types. The tabular view also gives access to additional details about each virus (e.g. common name of the host, associated publications, and GenBank accession numbers). Tables can be sorted by clicking on the column headers in the tables. Information in the tables and the tree are hyperlinked appropriately. Tables can be downloaded as comma separated (CSV) text files.

Figure 19. *Reference Genomes*

## 6.2. VARIANT GENOMES

The *Variant Genomes* tab links to PV variant nomenclature rules, enables browsing variants with curated details in a table and searching for to all human and animal variant sequences at the NCBI site.

### **6.3. TRANSCRIPT MAPS**

The *Transcript Maps* page provides links to description of PV transcripts with graphical representations.

### **6.4. PROTEINS**

The *Proteins* page contains links to a page that contains links to a) a table with all PV-related PDB structures or to the search interface which allows users to search for specific PV-related PDB structures in PaVE and b) a graphical representation of PV proteins.

### **6.5. EPIGENETICS**

The *Epigenetics* page contains links to epigenetic data visualized in a genome browser.

### **6.6. TAXONOMY**

The *Taxonomy* page contains links to the Taxonomy Concept page which describes the PV classification rules and the *Submission* page which provides current procedure for designation of a new viral type (*Submission*).

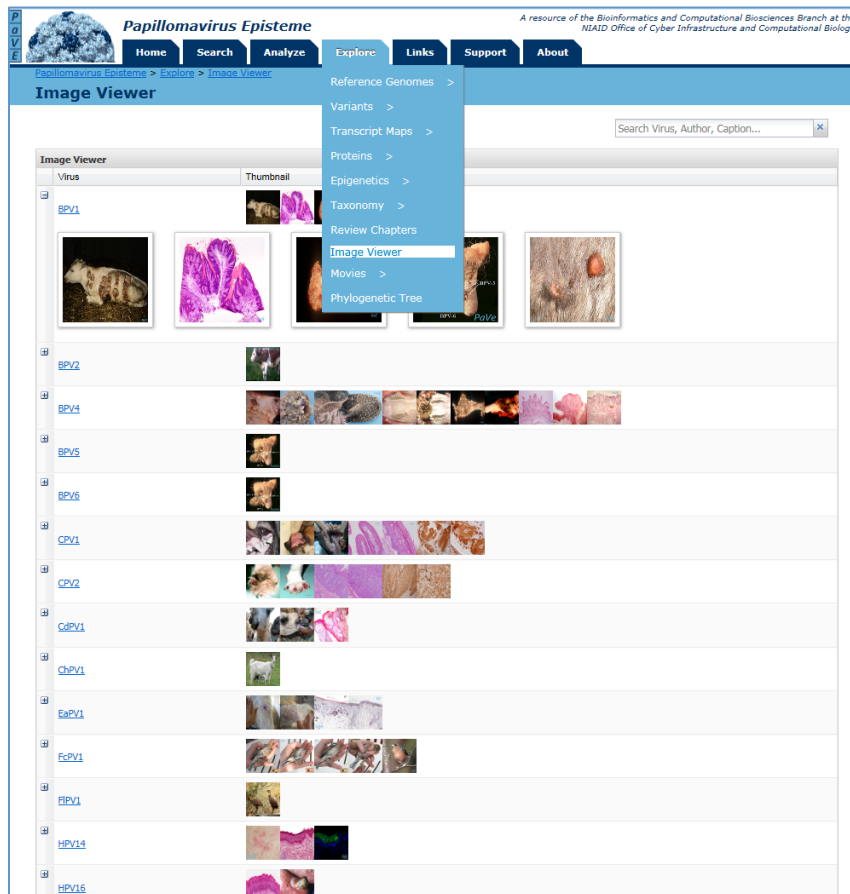
### **6.7. REVIEW CHAPTERS**

The *Review Chapter* page includes links to chapters in the Special Issue of Virology describing the functional genomics of Papillomaviruses. Each article in the Special Issue provides an in-depth, unbiased and encyclopedic overview of different aspects of papillomavirus genomics and proteomics, written by experts in the papillomavirus field. The main aim of each article is to completely document details about viral proteins, cis-elements, evolution, and disease. This Special Issue is intended to be a companion resource to the PaVE website and will serve as an important reference resource for the scientific community. You can visit the blog site for this special issue at <http://www.virologyhighlights.com/?p=144>.

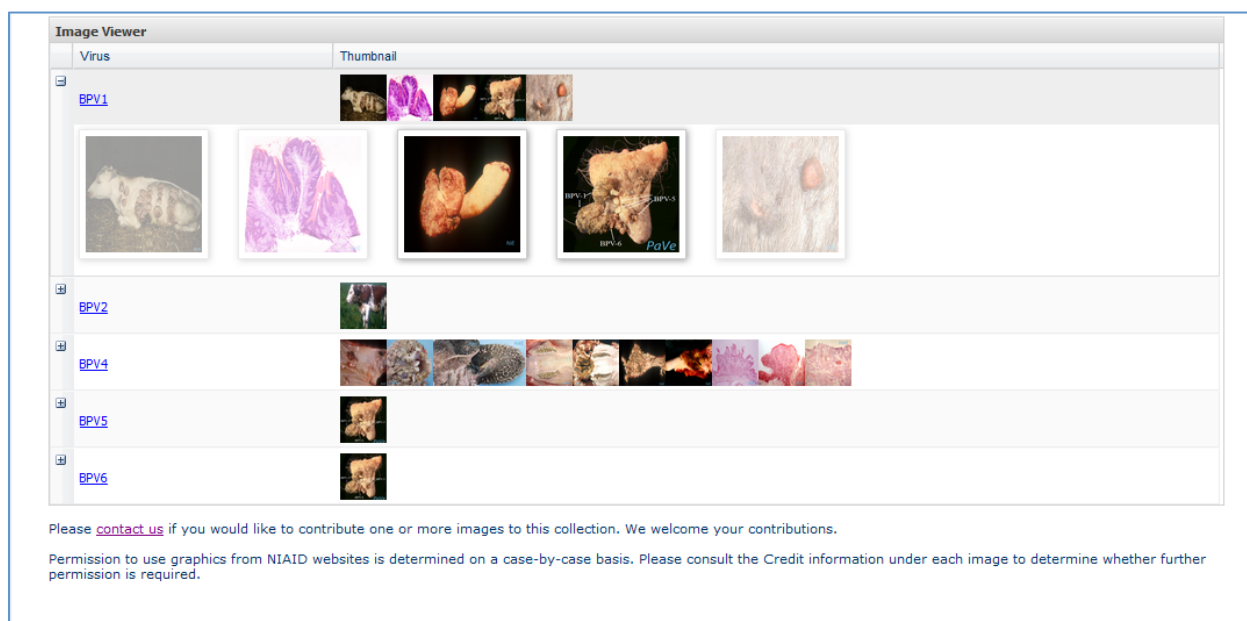
### **6.8. IMAGE VIEWER**

The *Image* tab in the top navigation bar links to a table of thumbnail images of viral (warts, lesions) grouped by PV type (Figure 21a). To filter by virus type, users can click on the triangle next to the column title and enter the term in the field available to filter by (figure 21a). Users can also search by the virus type, contributor, and by terms within the description by entering the term in the Search box. Search is PV type specific, in that results retrieved for contributor or description terms will return all images (including those that don't match the contributor or description terms) for PV types for which one or more images that match the criteria are available in PaVE (figure 21 b). However, images that don't match the criteria are greyed out. Certain images may be associated with more than one PV type.

Clicking on an image displays the image along with the caption, which includes a description of the image (tissue, organ, pathology etc.), the contributor name and relevant copyright information (filter 21c).



**Figure 21a. Images grouped by PV types.**



**Figure 21b. Images retrieved based on author search. Images contributed by the searched author are highlighted.**



## 7. SEQUENCE AND STRUCTURE VIEWERS

### 7.1. LOCUS VIEWER

The locus viewer page (figure 23a and b) displays graphical linear and circular representations of the viral genome with annotated features. Genes, regions and proteins can only be viewed in the linear view. Each feature can be selected, and details for that feature will be displayed in a panel below. Details are tailored to the type of sequence (Genome/Gene/Region/Protein) being viewed as described below. Users can download sequences in alternative formats, namely FastA, GenBank or EMBL and also link to the original submitter sequence at the NCBI site.

#### a) Genome/Gene/Region

1. The nucleotide sequence of the genome or gene or region is displayed below this feature panel
2. The sequence of the selected feature is highlighted in the nucleotide sequence of the genome and if the selected feature is a gene, the corresponding protein sequence is displayed in the Selected Feature Details box. The protein sequence can be aligned to other PaVE proteins via BLAST and the results copied and /or downloaded.

#### b) Protein

When viewing a protein sequence, users can link to available 3D structures for the protein in the Structure Viewer (Figure 23; Structure Viewer).

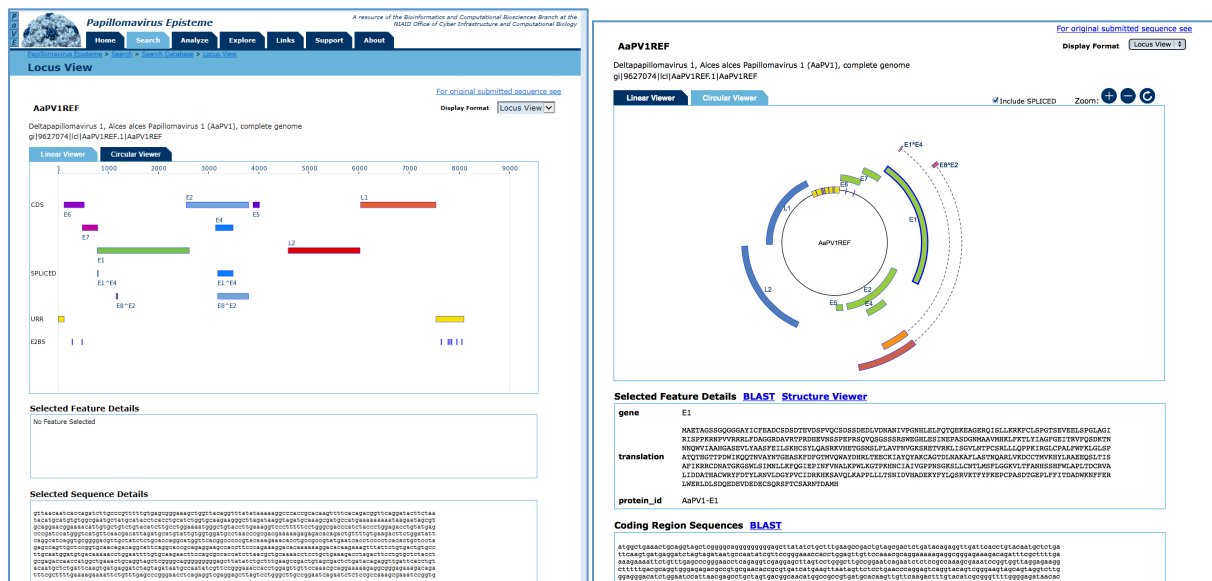


Figure 23a. Locus Viewer, Linear and Circular Displays (Genome)



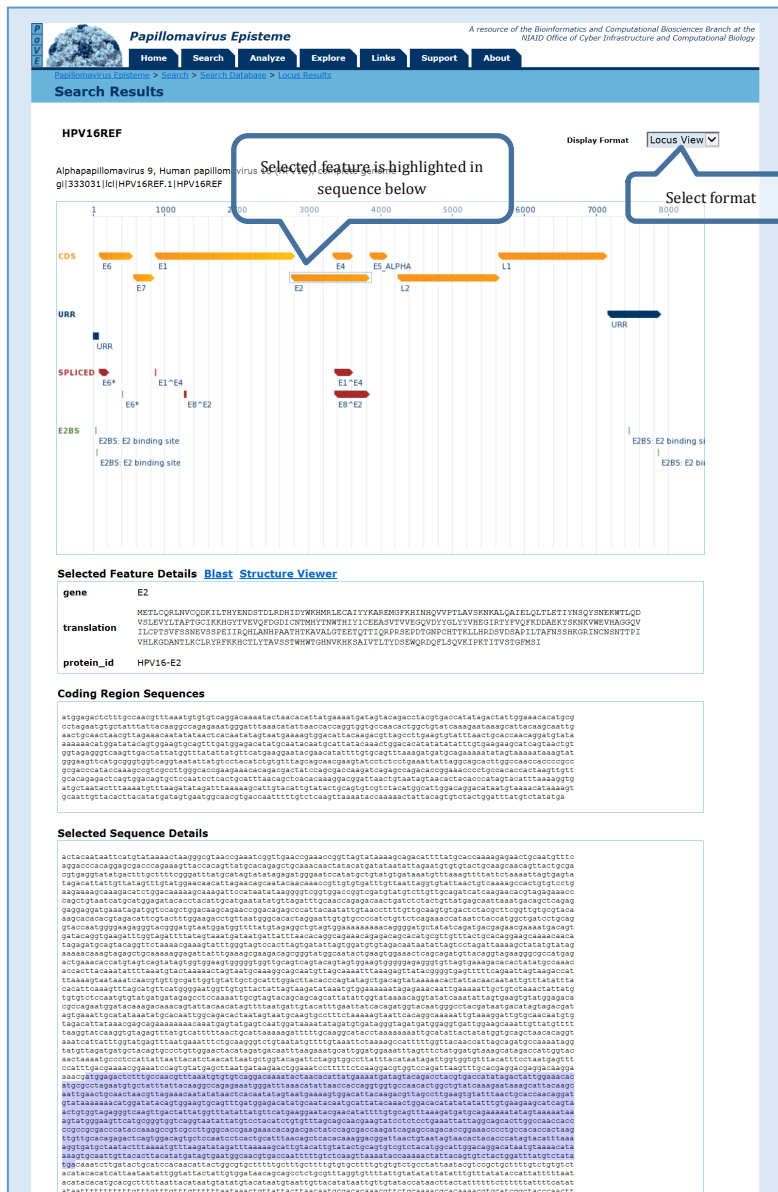


Figure 23b. Locus Viewer.

## 7.2. STRUCTURE VIEWER

The Structure Viewer page (figure 24) displays experimentally derived papillomavirus protein structures retrieved from the Protein Data Bank (PDB; <http://www.rcsb.org>) and integrated in PaVE. The features are listed below.

1. Structures are visualized using JSMol, which provides users the ability to zoom, view specific chains, alter the display of the structure, alter the color and background etc. Pairwise sequence comparisons between the solved structure and any homologous protein in the PaVE database are computed and the results displayed in a table. Users can also save and download the images by Right Clicking the image and choosing these options under the File tab.



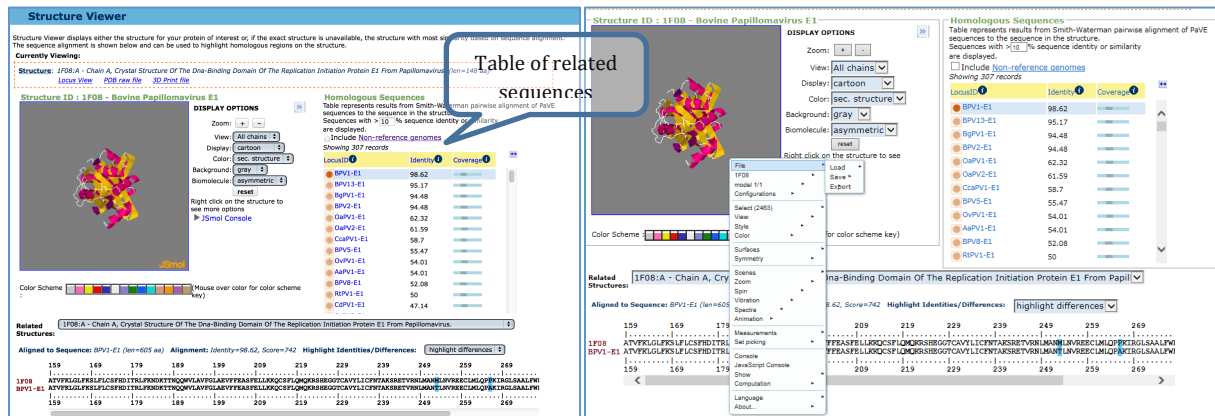


Figure 24. Structure Viewer.

- Click on any of the homologous proteins (BLASTP E<1e-5) to view the resulting alignment. Thus if a structure has been solved for any one papillomavirus protein or domain (e.g. E2), this tool allows the user to directly predict the structural location of selected sequences from other homologous papillomavirus proteins.
- The user can opt to color code differences and identities between the sequences and can use the mouse to select and highlight portions of the sequence alignment, which are then displayed on the structure.
- Users can also link to the locus view of the structural motifs, the PDB raw file and 3D print files on the [NIH 3D Print exchange](#).

## 8. FREQUENTLY ASKED QUESTIONS (FAQS)

### 1. How do I use the keyword search feature?

The keyword search feature uses a simple syntax. Add a "+" before a term (such as: +feline) to require it be present in the search results. Add a "-" before a term (such as: -ovine) to require it not be present in the results. Append an asterisk to the end of a term to act as a wildcard (such as: HPV16\*) to include all terms that begin with the term. Use quotes to search for complete phrases (such as: "complete genome").

### 2. How do I search the PaVE database by Taxonomy?

PaVE features a taxonomy search function. To search based on taxonomy click on the "Search" tab in the PaVE web site top navigation. Under "Keyword Search" open the pull-down menu next to "More Options" to reveal the advanced search options including the taxonomy search. With the options maximized, you will see "Taxonomy" listed in the lower right hand corner of options. Click on the drop down menu next to "any" under "Taxonomy" then click "Select from list". Your browser will then display a window for making your selections. To select a taxonomic genus, including all of the subgroups below it, simply click on the check box next to the genus name. You may select multiple genera in this way. Each selection that you make will appear in the "selected ids" section at the top of the window. Scroll up and down to see all of the groups. To make finer selections at the species level, expand the taxonomic genus view by clicking on the "+" icon next to the genus name. You may then select or deselect individual species. When all selections have been made. Click the "Done" button at the top of the window. The search filter will instantly be applied to the search results. The Taxonomy filter will display "..." instead of "any" to indicate that a filter has been applied.

### 3. How do I specify a "homology cutoff value" for my BLAST results?

Use the E-value cutoff function in PaVE's BLAST search window. Click on the "Search" tab in the PaVE web site top navigation, then switch from "Keyword Search" to "BLAST Search" by clicking on the BLAST Search tab. Under the BLAST Search tab, you will see the E-value cutoff option. E-values are the probability of finding a similar hit by

chance alone, thus the lower the E-value, the higher the quality of your BLAST hit. By default, the value is set to "10" which will return many results. To filter at a higher stringency, set a lower E- value. Cutoff E-value options are: 10, 0.1, 1e-05, and 1e-10.

**4. I've submitted a BLAST query and I see a list of results, how do I view the BLAST sequence hits?**




BLAST results on PaVE are returned in a table format sorted by E-value. To view the sequence query aligned to the BLAST hit, click on the triangle icon on the left hand side of the results table to drop down the alignment results.

**5. When I am observing a sequence record in Locus View, how can I display the sequence of a particular feature?**

Click on your feature of interest (coding sequence, origin of replication, source, etc.). The pane "Selected Feature Details" will display detailed information about the feature that you have selected. The pane "Selected Sequence Details" will display the entire locus sequence with your selected feature highlighted.

**6. How does the color scheme of the icons used to indicate availability of protein structures relate to the sequence homology computation**

This is determined by the alignment of protein sequence with sequence used in the PDB structure and is explained below:

-  True structure available
-  Closely resembling structure where alignment is >10%
-  No available structure with alignment >10%

**7. What are the minimum browser requirements to use PaVE?**

PaVE has been tested on IE v10 and 11, Safari 6 and higher and Firefox 23 and higher. JavaScript must be enabled in your browser's settings. Please note that the visualization of the multiple sequence alignments in Jalview is not compatible with very recent versions of Firefox. Please keep your java version up-to-date (<http://www.oracle.com/technetwork/java/javase/downloads/index.html>).